KOMPARASI *NAIVE BAYES DAN SUPPORT VECTOR MACHINE* DALAM MENDETEKSI SARKASME TEKS TWITTER (X)

Syifa Azzahra[1], Bayu Sulistiyanto Ipung Sutejo[2], Diah Aryani[3], Jefry Sunupurwa Asri[4]
Program Studi Teknik Informatika, Fakultas Ilmu Komputer
Universitas Esa Unggul[1,2,3,4]
Jalan Arjuna Utara No 9, Kebon Jeruk, Jakarta Barat
syfazzahra14@student.esaunggul.ac.id[1], bayu.sulistiyanto@esaunggul.ac.id[2],
diah.aryani@esaunggul.ac.id[3], jefry.sunupurwa@esaunggul.ac.id[4]

Abstrak

Penggunaan bahasa tidak baku pada media sosial menjadi salah satu tantangan dalam mendeteksi sarkasme karena makna yang disampaikan tidak selalu dapat dikenali hanya berdasarkan kata-kata yang terdapat dalam teks. Penelitian bertujuan membandingkan kinerja *Naive Bayes* dan *Support Vector Machine* (SVM) dalam mendeteksi sarkasme pada teks Twitter (X) berbahasa Indonesia tidak baku. *Wilcoxon Signed-Rank Test* digunakan untuk menguji signifikansi perbedaan kinerja kedua algoritma. Hasil audit menghasilkan 15.072 tweet unik yang terdiri atas 12.057 data latih dan 3.015 data uji. Hasil pengujian menunjukkan bahwa *Naive Bayes* memperoleh *accuracy* 69,45%, *precision* 67,66%, *recall* 64,19%, dan *F1-Score* 65,88%, sedangkan SVM memperoleh *accuracy* 68,39%, *precision* 66,51%, *recall* 62,82%, dan *F1-Score* 64,61%. Pada *5-Fold Cross Validation*, *Naive Bayes* menghasilkan rata-rata *accuracy* sebesar 68,74% dengan standar deviasi ±0,33%, sedangkan SVM sebesar 68,17% dengan standar deviasi ±1,14%. Hasil *Wilcoxon Signed-Rank Test* menghasilkan *p-value* sebesar 0,4375 (>0,05), yang menunjukkan bahwa perbedaan kinerja kedua algoritma tidak signifikan secara statistik. Dengan demikian, *Naive Bayes* menunjukkan hasil evaluasi yang sedikit lebih tinggi secara numerik, tetapi belum dapat dinyatakan memiliki kinerja yang secara signifikan lebih unggul dibandingkan SVM pada dataset penelitian ini.

Kata Kunci: deteksi sarkasme, Twitter (X), *Naive Bayes, Support Vector Machine, TF-IDF*

Abstract

The use of non-standard language on social media has become one of the challenges in detecting sarcasm because the meaning conveyed is not always recognizable based solely on the words present in the text. This study aims to compare the performance of Naive Bayes and Support Vector Machine (SVM) in detecting sarcasm in non-standard Indonesian Twitter text (X). The Wilcoxon Signed-Rank Test is used to test the significance of the performance difference between the two algorithms. The audit results in 15,072 unique tweets, consisting of 12,057 training data and 3,015 test data. The test results show that Naive Bayes achieved an accuracy of 69.45%, precision of 67.66%, recall of 64.19%, and F1-Score of 65.88%, while SVM achieved an accuracy of 68.39%, precision of 66.51%, recall of 62.82%, and F1-Score of 64.61%. In the 5-fold cross-validation, Naive Bayes produced an average accuracy of 68.74% with a standard deviation of ±0.33%, while SVM achieved 68.17% with a standard deviation of ±1.14%. The Wilcoxon Signed-Rank Test results in a p-value of 0.4375 (>0.05), indicating that the performance difference between the two algorithms is not statistically significant. Thus, Naive Bayes shows slightly higher evaluation results numerically, but it cannot yet be stated that it has significantly superior performance compared to SVM on this research dataset.

Keywords*: sarcasm detection, Twitter (X), Naive Bayes, Support Vector Machine, TF-IDF.*

PENDAHULUAN

Media sosial menjadi salah satu ruang komunikasi digital yang memungkinkan pengguna menyampaikan opini, kritik, dan berbagai bentuk ekspresi secara cepat. Twitter (X) merupakan salah satu sumber data yang banyak dimanfaatkan dalam penelitian *Natural Language Processing* (NLP) karena menyediakan teks yang singkat, beragam, dan bersifat informal. Karakteristik tersebut menjadikan Twitter (X) relevan untuk mengkaji berbagai fenomena kebahasaan, termasuk penggunaan bahasa Indonesia tidak baku dan sarkasme dalam komunikasi digital. Penggunaan bahasa pada Twitter (X) berbeda dari bahasa formal karena banyak ditemukan singkatan, variasi ejaan, istilah *slang*, dan bentuk penulisan informal. Kondisi tersebut menjadi semakin kompleks ketika teks mengandung sarkasme, karena makna yang disampaikan dapat berbeda atau berlawanan dengan makna literalnya. Deteksi sarkasme tidak hanya bergantung pada

kata yang muncul dalam teks, tetapi juga berkaitan dengan konteks dan maksud yang ingin disampaikan. Ketidaksesuaian antara makna literal dan maksud yang disampaikan menjadi salah satu faktor yang menyebabkan deteksi sarkasme lebih kompleks dalam pengolahan bahasa alami (Alqahtani et al., 2023). Penggunaan bahasa pada Twitter (X) berbeda dari bahasa formal karena banyak ditemukan singkatan, variasi ejaan, istilah *slang*, dan bentuk penulisan informal. Kondisi tersebut menjadi semakin kompleks ketika teks mengandung sarkasme, karena makna yang disampaikan dapat berbeda atau berlawanan dengan makna literalnya. Deteksi sarkasme tidak hanya bergantung pada kata yang muncul dalam teks, tetapi juga berkaitan dengan konteks dan maksud yang ingin disampaikan. Ketidaksesuaian antara makna literal dan maksud yang disampaikan menjadi salah satu faktor yang menyebabkan deteksi sarkasme lebih kompleks dalam pengolahan bahasa alami (Alqahtani et al., 2023). Oleh karena itu, diperlukan pendekatan yang mampu mempelajari pola penggunaan kata dalam teks secara sistematis untuk membantu mengidentifikasi karakteristik sarkastik tersebut. Salah satu pendekatan yang dapat digunakan adalah *machine learning*, yang memungkinkan proses klasifikasi teks dilakukan berdasarkan pola yang dipelajari dari data.

Dalam klasifikasi teks, *Naive Bayes* dan *Support Vector Machine* (SVM) merupakan dua algoritma yang memiliki pendekatan berbeda dalam mengenali pola antarkelas. *Naive Bayes* menggunakan pendekatan probabilistik untuk menghitung kemungkinan suatu teks termasuk ke dalam kelas tertentu, sedangkan SVM menentukan batas pemisah antarkelas pada ruang fitur. Keduanya dapat digunakan bersama *Term Frequency-Inverse Document Frequency* (TF-IDF) untuk mengubah teks menjadi representasi numerik sebelum dilakukan proses klasifikasi. Penelitian terdahulu menunjukkan bahwa kinerja kedua algoritma tersebut dapat berbeda bergantung pada karakteristik data dan permasalahan klasifikasi yang digunakan. Khotimah dan Utami (2022) menunjukkan bahwa *Naive Bayes* memperoleh kinerja yang lebih tinggi pada data Twitter, sementara penelitian lain menunjukkan bahwa *SVM* dapat memberikan hasil yang lebih baik pada klasifikasi teks media sosial. (Khotimah & Utami, 2022; Panjaitan et al., 2025)

Berdasarkan kondisi tersebut, penelitian ini memberikan kontribusi melalui perbandingan langsung *Naive Bayes* dan *SVM* pada teks Twitter (X) berbahasa Indonesia tidak baku yang mengandung sarkasme. Kedua algoritma diberikan perlakuan yang sama, meliputi dataset, prapemrosesan, representasi fitur TF-IDF, pembagian data, dan prosedur evaluasi. Selain membandingkan *accuracy*, *precision*, *recall*, dan *F1-Score*, penelitian ini menggunakan *5-Fold Cross Validation* untuk melihat konsistensi kinerja serta *Wilcoxon Signed-Rank Test* untuk mengetahui apakah perbedaan performa kedua algoritma signifikan secara statistik. Pendekatan tersebut memungkinkan perbandingan dilakukan tidak hanya berdasarkan nilai metrik evaluasi, tetapi juga berdasarkan konsistensi dan signifikansi perbedaan kinerja kedua algoritma.

Penelitian ini bertujuan menganalisis dan membandingkan kinerja *Naive Bayes* dan *SVM* dalam mendeteksi sarkasme pada teks Twitter (X) berbahasa Indonesia tidak baku serta mengetahui signifikansi perbedaan kinerja kedua algoritma. Penelitian dibatasi pada klasifikasi biner sarkasme dan non-sarkasme menggunakan dataset dari *Hugging Face* yang telah melalui proses audit dan menghasilkan 15.072 tweet unik. Prapemrosesan dilakukan melalui *case folding*, penghapusan URL, *mention*, *hashtag*, angka, dan tanda baca, tokenisasi, normalisasi kata tidak baku, *stopword removal*, serta *stemming* menggunakan Sastrawi. Representasi teks menggunakan TF-IDF dengan konfigurasi *max_features*=5000, *ngram_range*=(1,1), *min_df*=2, dan *max_df*=0,95.

PENELITIAN RELEVAN

Berdasarkan kajian terhadap penelitian terdahulu (Tabel 1), masih terdapat keterbatasan penelitian yang secara khusus membandingkan *Naive Bayes* dan *Support Vector Machine* dalam mendeteksi sarkasme pada teks Twitter (X) berbahasa Indonesia tidak baku dengan penggunaan representasi fitur dan kondisi pengujian yang setara. Selain itu, penelitian yang mengintegrasikan perbandingan kedua algoritma tersebut dengan *Wilcoxon Signed-Rank Test* untuk menguji signifikansi perbedaan kinerja juga masih terbatas. Oleh karena itu, penelitian ini dilakukan untuk membandingkan kinerja *Naive Bayes* dan *Support Vector Machine* secara langsung serta mengetahui apakah perbedaan performa yang dihasilkan memiliki signifikansi secara statistik.

Tabel 1. Analisis Penelitian Relevan

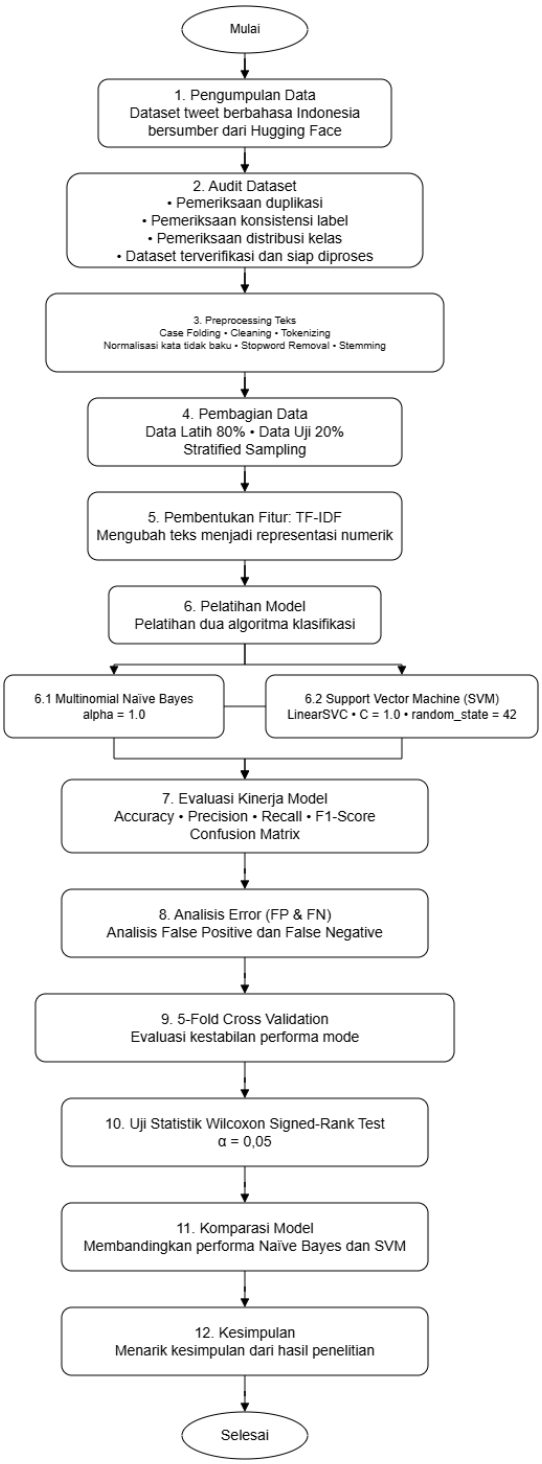
Penulis	Metode	*Dataset*	Hasil Utama	Kekurangan
Rosid et al. (2022)	*SVM, Random Forest, KNN*	Twitter berbahasa Indonesia	Membandingkan beberapa algoritma	Belum membandingkan *Naive Bayes* dan *SVM*

			machine learning untuk deteksi sarkasme	sebagai dua model utama dengan kondisi pengujian yang sama
Khotimah & Utami (2022)	*Naive Bayes, KNN, SVM, TF-IDF*	Data Twitter	*Naive Bayes* memperoleh kinerja lebih tinggi pada dataset yang digunakan	Fokus pada klasifikasi data Twitter secara umum, bukan khusus deteksi sarkasme
Lestari & Hutagalung (2025)	*Logistic Regression, Naive Bayes, SVM, TF-IDF*	Teks berbahasa Indonesia	*TF-IDF* efektif sebagai representasi fitur untuk klasifikasi teks	*TF-IDF* memiliki keterbatasan dalam menangkap konteks semantik kompleks seperti sarkasme
Panjaitan et al. (2025)	*SVM, Naive Bayes*, dan algoritma machine learning lainnya	Data media sosial	*SVM* memperoleh accuracy rata-rata lebih tinggi dibandingkan *Naive Bayes*	Objek penelitian berupa analisis sentimen, bukan deteksi sarkasme pada teks Indonesia tidak baku
Alqahtani et al. (2023)	*Systematic Literature Review*	Berbagai penelitian deteksi sarkasme pada media sosial	Twitter merupakan salah satu sumber data yang banyak digunakan dan pendekatan deep learning semakin berkembang	Tidak secara khusus membandingkan *NB* dan *SVM* pada teks Twitter Indonesia tidak baku

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen untuk membandingkan kinerja *Naive Bayes* dan *Support Vector Machine (SVM)* dalam mendeteksi sarkasme pada teks Twitter (X) berbahasa Indonesia tidak baku. Penelitian menerapkan *supervised learning* karena data memiliki label sarkasme dan non-sarkasme pada atribut *is_sarcasm*. Dataset diperoleh dari *Hugging Face* melalui *adealvii/indonesian_tweet_balances_12k* dan *adealvii/sarcasm-indo-crawl-9k*. Kedua dataset digabungkan menjadi 21.784 tweet, kemudian dilakukan audit duplikasi berdasarkan atribut *full_text* sehingga diperoleh 15.072 tweet unik. Prapemrosesan dilakukan melalui *case folding*, penghapusan URL, *mention*, hashtag, angka, dan tanda baca, tokenisasi, normalisasi kata tidak baku menggunakan kamus slang dan baku, *stopword removal*, serta *stemming* dengan *Sastrawi*. Data kemudian dibagi menggunakan *train-test split* 80:20, menghasilkan 12.057 data latih dan 3.015 data uji. Selanjutnya, teks direpresentasikan menggunakan *TF-IDF* dengan konfigurasi *max_features*=5000, *ngram_range*=(1,1), *min_df*=2, dan *max_df*=0,95.

Naive Bayes dan *SVM* dilatih menggunakan data dengan representasi fitur dan perlakuan yang sama agar perbandingan kinerja dilakukan secara adil. Evaluasi model menggunakan *accuracy*, *precision*, *recall*, *F1-Score*, dan *confusion matrix* untuk mengukur kemampuan klasifikasi sarkasme dan non-sarkasme. Selain itu, dilakukan *5-Fold Cross Validation* untuk melihat konsistensi performa kedua model pada beberapa pembagian data. Hasil *5-Fold Cross Validation* kemudian dianalisis menggunakan *Wilcoxon Signed-Rank Test* untuk mengetahui apakah terdapat perbedaan kinerja yang signifikan antara *Naive Bayes* dan *SVM*. Dengan demikian, perbandingan tidak hanya didasarkan pada nilai metrik evaluasi, tetapi juga mempertimbangkan konsistensi dan signifikansi perbedaan kinerja kedua algoritma.



Gambar 1. Alur Penelitian

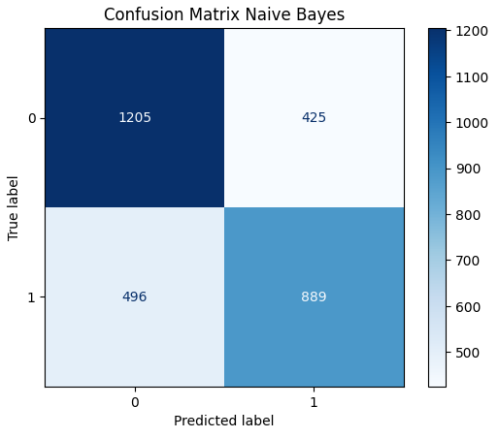
HASIL DAN PEMBAHASAN

Proses audit dataset dilakukan untuk memastikan kualitas dan keunikan data sebelum masuk ke tahap pemodelan. Pemeriksaan duplikasi dilakukan berdasarkan atribut *full_text* dari total 21.784 tweet yang berasal dari dua sumber dataset. Hasil pemeriksaan ditunjukkan pada Tabel 2, menunjukkan terdapat 6.712 data yang memiliki teks duplikat. Setelah data duplikat dihapus, tersisa 15.072 tweet unik yang digunakan untuk tahap prapemrosesan dan pemodelan. Dengan demikian, proses audit mengurangi sekitar 30,81% dari keseluruhan data awal, sehingga dataset yang digunakan lebih terkontrol dan terhindar dari pengulangan teks yang sama.

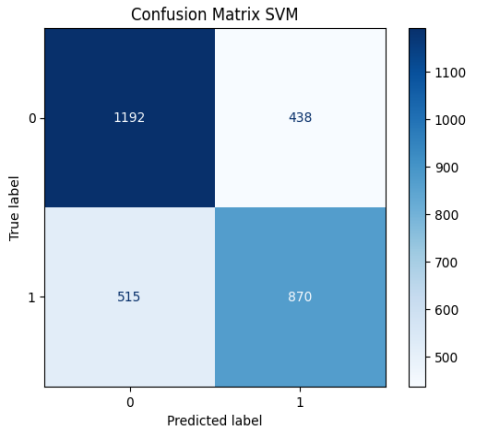
Tabel 2. Hasil Audit Dataset

Tahapan	Jumlah Tweet
Dataset awal	21.784
Data duplikat dihapus	6.712
Dataset setelah audit	15.072

Tahapan preprocessing berhasil menyeragamkan variasi penulisan tidak baku pada teks Twitter (X), misalnya kata "yg" menjadi "yang", "br" menjadi "baru", dan "udh" menjadi "sudah" setelah proses normalisasi, serta kata berimbuhan seperti "menjabat" menjadi bentuk dasar "jabat" setelah proses stemming. Hasil transformasi *TF-IDF* menghasilkan matriks fitur berukuran 12.057×5.000 pada data latih dan 3.015×5.000 pada data uji, yang selanjutnya digunakan sebagai masukan pada proses pelatihan dan pengujian model.



Gambar 2. *Confusion matrix Naive Bayes*



Gambar 3. *Confusion matrix Support Vector Machine*

Confusion matrix masing-masing model ditunjukkan pada Gambar 2 dan Gambar 3 dimana hasil pengujian terhadap 3.015 data uji menunjukkan bahwa *Naive Bayes* menghasilkan 1.205 *True Negative* (TN), 425 *False Positive* (FP), 496 *False Negative* (FN), dan 889 *True Positive* (TP). Dengan hasil tersebut, sebanyak 2.094 data berhasil diklasifikasikan dengan benar, sedangkan 921 data mengalami kesalahan klasifikasi. Sementara itu, SVM menghasilkan 1.192 TN, 438 FP, 515 FN, dan 870 TP, sehingga terdapat 2.062 prediksi yang benar dan 953 prediksi yang salah. Hasil ini menunjukkan bahwa *Naive Bayes* menghasilkan prediksi benar yang sedikit lebih banyak dibandingkan SVM pada data uji. Temuan tersebut sejalan dengan Khotimah dan Utami (2022) yang menunjukkan bahwa kinerja *Naive Bayes* dan SVM pada data Twitter dapat berbeda bergantung pada karakteristik data dan proses klasifikasi yang digunakan.

Berdasarkan metrik evaluasi (Tabel 2), *Naive Bayes* memperoleh *accuracy* sebesar 69,45%, *precision* 67,66%, *recall* 64,19%, dan *F1-Score* 65,88%. Sementara itu, SVM memperoleh *accuracy* sebesar 68,39%, *precision* 66,51%, *recall* 62,82%, dan *F1-Score* 64,61%. Dengan demikian, *Naive Bayes* memiliki nilai sedikit lebih tinggi pada seluruh metrik, dengan selisih *accuracy* sebesar 1,06 poin persentase dan *F1-Score* sebesar 1,27 poin persentase. Hasil tersebut sejalan dengan penelitian Rifa'i dan Mukti (2025) yang menunjukkan bahwa *Naive Bayes* dapat digunakan sebagai *baseline* yang cukup baik pada klasifikasi teks dengan representasi TF-IDF. Penggunaan TF-IDF juga memungkinkan fitur kata direpresentasikan berdasarkan tingkat kepentingannya dalam dokumen sehingga dapat digunakan sebagai masukan bagi algoritma klasifikasi teks (Gerliandeva et al., 2024)

Tabel 3. Hasil Evaluasi Kinerja Algoritma

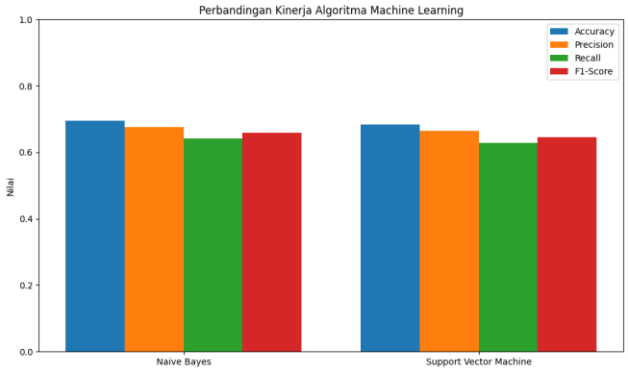
Algoritma	*Accuracy*	*Precision*	*Recall*	*F1-Score*
Naive Bayes	69,45%	67,66%	64,19%	65,88%
Support Vector Machine	68,39%	66,51%	62,82%	64,61%

Pada *confusion matrix*, kedua model menghasilkan jumlah FN yang lebih besar dibandingkan FP. *Naive Bayes* menghasilkan 425 FP dan 496 FN, sedangkan SVM menghasilkan 438 FP dan 515 FN.

Artinya, kedua model lebih sering gagal mengenali tweet yang sebenarnya mengandung sarkasme dibandingkan salah mengategorikan tweet non-sarkasme sebagai sarkasme. Kondisi ini berkaitan dengan karakteristik sarkasme yang sering disampaikan secara tidak langsung dan dapat memiliki makna yang berbeda dari makna literalnya. Deteksi sarkasme pada teks media sosial memang menjadi lebih kompleks ketika model hanya mengandalkan informasi yang terdapat pada teks tanpa mempertimbangkan konteks penggunaan bahasa (Alqahtani et al., 2023; Yunitasari et al., 2019).

Analisis kesalahan klasifikasi kemudian dilakukan dengan melihat contoh *False Positive* dan *False Negative* pada *Naive Bayes*. Tweet "dari dulu indonesia emang ga pernah baik baik aja bos" merupakan non-sarkasme, tetapi diprediksi sebagai sarkasme. Penggunaan ungkapan "ga pernah baik baik aja" dan kata "bos" dapat menyerupai pola bahasa yang bernada sindiran. Sebaliknya, tweet "keren bgt sholat ied pake ai ghibli ya ini" merupakan sarkasme, tetapi diklasifikasikan sebagai non-sarkasme. Penggunaan kata "keren bgt" secara literal memberikan kesan positif sehingga konteks sindiran pada tweet tersebut lebih sulit dikenali. Pola seperti ini menunjukkan bahwa kata yang memiliki makna positif secara literal belum tentu menunjukkan maksud positif dalam sebuah tweet (Alqahtani et al., 2023; Rosid et al., 2022)

Kesalahan tersebut memperlihatkan bahwa klasifikasi sarkasme tidak hanya bergantung pada keberadaan kata tertentu, tetapi juga pada hubungan kata dan konteks penggunaannya. TF-IDF dapat memberikan representasi yang baik untuk klasifikasi teks, tetapi pembobotannya tetap berfokus pada karakteristik kata dalam dokumen sehingga belum sepenuhnya menangkap makna kontekstual dan maksud tersirat. Keterbatasan tersebut juga terlihat pada penelitian Lestari dan Hutagalung (2025) yang menunjukkan bahwa TF-IDF efektif digunakan dalam klasifikasi teks berbahasa Indonesia, tetapi masih memiliki keterbatasan ketika berhadapan dengan informasi semantik yang lebih kompleks. Pendekatan yang lebih berorientasi pada konteks seperti model berbasis bahasa dapat menjadi alternatif untuk menangkap karakteristik tersebut (Jakarta et al., n.d.; Perdana et al., 2026)



Gambar 4. Perbandingan Kinerja *Naïve Bayes* dan SVM

Disini *Naive Bayes* menunjukkan kinerja yang sedikit lebih tinggi dibandingkan *SVM* pada seluruh metrik evaluasi. Selisih *accuracy* sebesar 1,06 poin persentase menunjukkan bahwa perbedaan kinerja kedua model relatif kecil. Hasil ini sejalan dengan penelitian Rifa'i dan Mukti (2025) yang menunjukkan bahwa *Naive Bayes* dapat memberikan kinerja yang konsisten pada klasifikasi teks berdimensi tinggi. Kondisi tersebut berkaitan dengan karakteristik fitur *TF-IDF* yang memiliki dimensi tinggi dan bersifat *sparse*. *Naive Bayes* memanfaatkan probabilitas kemunculan fitur pada masing-masing kelas dalam menentukan hasil klasifikasi, sehingga dapat bekerja dengan baik pada data teks dengan jumlah fitur yang besar dan distribusi kata yang beragam (Rifa'i & Mukti, 2025).

Sementara itu, *SVM* dalam penelitian ini menggunakan *LinearSVC* dengan nilai C=1,0. Model tersebut membentuk batas pemisah linear untuk membedakan kelas pada ruang fitur. Pada data sarkasme, pemisahan antarkelas dapat menjadi lebih sulit ketika pola bahasa yang digunakan saling tumpang tindih. Sarkasme tidak selalu ditunjukkan melalui kata tertentu karena kata yang secara literal memiliki makna positif dapat digunakan untuk menyampaikan sindiran. Kondisi tersebut membuat pola sarkasme tidak selalu dapat dipisahkan hanya berdasarkan kemunculan kata dalam

teks. Hasil ini berbeda dengan penelitian Panjaitan et al. (2025) dan Yusupa dan Tarigan (2024) yang menunjukkan kinerja *SVM* lebih tinggi pada klasifikasi teks media sosial. Perbedaan hasil tersebut menunjukkan bahwa kinerja algoritma dapat dipengaruhi oleh karakteristik dataset dan permasalahan klasifikasi yang digunakan (Panjaitan et al., 2025; Yusupa & Tarigan, n.d.)

Dalam penelitian ini, selisih kinerja *Naive Bayes* dan *SVM* yang relatif kecil lebih menggambarkan kondisi dataset dan representasi *TF-IDF* yang digunakan. Temuan tersebut juga memperlihatkan bahwa algoritma yang memiliki kinerja lebih tinggi pada satu penelitian belum tentu menghasilkan kinerja yang sama pada dataset dengan karakteristik berbeda. Dengan kata lain, hasil penelitian ini tidak menunjukkan bahwa *Naive Bayes* selalu lebih unggul, tetapi menunjukkan bahwa *Naive Bayes* memberikan hasil numerik yang sedikit lebih baik pada dataset sarkasme Twitter (X) berbahasa Indonesia tidak baku yang digunakan dalam penelitian ini.

Tabel 4. Hasil 5-Fold Cross Validation

Algoritma	Rata-rata *Accuracy*	Standar Deviasi
Naive Bayes	68,74%	±0,33%
Support Vector Machine	68,17%	±1,14%

Sebagai pengujian lanjutan untuk memastikan bahwa hasil klasifikasi tidak hanya bergantung pada satu pembagian data, dilakukan *5-Fold Cross Validation* (Tabel 4) menggunakan *StratifiedKFold* dengan lima *fold*. Pendekatan ini memberikan kesempatan setiap bagian data menjadi data validasi secara bergantian, sehingga performa model dapat diamati pada beberapa konfigurasi pembagian data dan proporsi kelas tetap terjaga pada setiap *fold* ((Prusty et al., 2022; Widodo et al., 2022) Hasil *5-Fold Cross Validation* menunjukkan bahwa *Naive Bayes* memperoleh rata-rata *accuracy* sebesar 68,74% dengan standar deviasi 0,33%, sedangkan SVM memperoleh rata-rata *accuracy* sebesar 68,17% dengan standar deviasi 1,14%. Selisih rata-rata *accuracy* kedua model hanya sebesar 0,57 poin persentase. Standar deviasi yang lebih kecil pada *Naive Bayes* menunjukkan bahwa nilai *accuracy* model tersebut memiliki variasi yang lebih rendah antar-fold dibandingkan *SVM*. Dengan demikian, *Naive Bayes* tidak hanya memiliki rata-rata *accuracy* yang sedikit lebih tinggi, tetapi juga menunjukkan performa yang lebih konsisten pada lima pembagian data yang digunakan.

Hasil *5-Fold Cross Validation* selanjutnya diuji menggunakan *Wilcoxon Signed-Rank Test* untuk mengetahui apakah perbedaan *accuracy* antara *Naive Bayes* dan *SVM* signifikan secara statistik (Tabel 5). Pengujian menghasilkan statistik *Wilcoxon* sebesar 4,0000 dan *p-value* sebesar 0,4375. Karena *p-value* lebih besar dari 0,05, perbedaan *accuracy* antara *Naive Bayes* dan *SVM* tidak signifikan secara statistik. Dengan kata lain, selisih performa yang muncul lebih mungkin mencerminkan variasi pada pembagian data daripada menunjukkan keunggulan algoritma tertentu.

Tabel 5. Hasil *Wilcoxon Signed-Rank Test*

Perbandingan	Statistik	*p-value*	α	Keputusan
Naive Bayes vs SVM	*4,0000*	*0,4375*	0,05	H_0 gagal ditolak

SIMPULAN

Berdasarkan hasil penelitian, *Naive Bayes* dan *Support Vector Machine* (SVM) menunjukkan kinerja yang relatif seimbang dalam mendeteksi sarkasme pada teks Twitter (X) berbahasa Indonesia tidak baku. *Naive Bayes* memperoleh *accuracy* sebesar 69,45%, *precision* 67,66%, *recall* 64,19%, dan *F1-Score* 65,88%, sedangkan *Support Vector Machine* memperoleh *accuracy* 68,39%, *precision* 66,51%, *recall* 62,82%, dan *F1-Score* 64,61%. Pada pengujian 5-Fold *Cross Validation*, *Naive Bayes* juga menghasilkan rata-rata *accuracy* sedikit lebih tinggi, yaitu 68,74% dibandingkan *Support Vector Machine* sebesar 68,17%. Namun, hasil *Wilcoxon Signed-Rank Test* menunjukkan *p-value* sebesar 0,4375 (>0,05), sehingga perbedaan kinerja kedua algoritma tidak terbukti signifikan secara statistik. Dengan demikian, *Naive Bayes* memiliki hasil numerik yang sedikit lebih baik, tetapi belum dapat dinyatakan secara signifikan lebih unggul dibandingkan dengan *Support Vector Machine*.
Hasil penelitian menunjukkan bahwa kemampuan klasifikasi sarkasme tidak hanya dipengaruhi oleh algoritma, tetapi juga oleh karakteristik data, tahapan prapemrosesan, dan representasi fitur yang

digunakan. Penelitian selanjutnya dapat mempertimbangkan pendekatan berbasis konteks seperti IndoBERT atau LSTM untuk menangkap makna implisit, ironi, dan ketidaksesuaian semantik yang belum sepenuhnya direpresentasikan oleh TF-IDF. Pengembangan penelitian juga dapat dilakukan dengan menambahkan anotasi tingkat sarkasme, menguji model pada platform media sosial lain seperti Instagram dan TikTok, serta menerapkan *hyperparameter tuning* yang lebih luas pada *Support Vector Machine* untuk mengetahui kemungkinan peningkatan kinerja model.

DAFTAR PUSTAKA

Alqahtani, A., Al-Henaki, L., & Alsheddi, A. (2023). Text-based Sarcasm Detection on Social Networks: A Systematic Review. *International Journal of Advanced Computer Science and Applications*, *14*(3). https://doi.org/10.14569/ijacsa.2023.0140336

Gerliandeva, A., Chrisnanto, Y., & Ashaury, H. (2024). Optimasi Klasifikasi Sentimen pada Komentar Online menggunakan Multinomial Naïve Bayes dan Ekstraksi Fitur TF-IDF serta N-grams. *Jurnal Pekommas*, *9*(2), 260–272. https://doi.org/10.56873/jpkm.v9i2.5585

Jakarta, U. N., Lestari, V. B., & Hutagalung, C. A. (n.d.). *J-KOMA Journal of Computer Science and Applications Evaluation of TF-IDF Extraction Techniques in Sentiment Analysis of Indonesian-Language Marketplaces Using SVM, Logistic Regression, and Naive Bayes*. 22–2025. https://doi.org/10.21009/j

Khotimah, A. C., & Utami, E. (2022). Comparison Naïve Bayes Classifier, K-Nearest Neighbor And Support Vector Machine In The Classification Of Individual On Twitter Account. *Jurnal Teknik Informatika (JUTIF)*, *3*(3). https://doi.org/10.20884/1.jutif.2022.3.3.254

Panjaitan, F., Ce, W., Oktafiandy, H., Kanugrahan, G., Ramdhani, Y., Hafizh, V., Putra, C., & Permai, A. (2025). Studi Komparatif Algoritma Machine Learning Pada Analisis Sentimen Media Sosial. In *Jurnal Mahasiswa Teknik Informatika)* (Vol. 9, Number 2).

Perdana, A., Sugandi, F., & Raihannicko, A. (2026). *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*.

Prusty, S., Patnaik, S., & Dash, S. K. (2022). SKCV: Stratified K-fold cross-validation on ML classifiers for predicting cervical cancer. *Frontiers in Nanotechnology*, *4*. https://doi.org/10.3389/fnano.2022.972421

Rifa'i, Z., & Mukti, B. P. (2025). Edu Komputika Journal Weakly Supervised Sentiment Analysis of Indonesian Rural Tourism Reviews: A TF-IDF Baseline for Melung Tourism Village Article Info. *Edu Komputika*, *12*(1). https://journal.unnes.ac.id/journals/edukom/

Rosid, M. A., Muharram, F., & Affandi, G. R. (2022). *Comparison of the Performance of Machine Learning Algorithms for Sarcasm Detection in Bahasa Perbandingan Kinerja Algoritma Machine Learning Untuk Mendeteksi Kalimat Sarkasme Dalam Bahasa Indonesia*. https://pssh.umsida.ac.id.

Widodo, S., Brawijaya, H., & Samudi, S. (2022). Stratified K-fold cross validation optimization on machine learning for prediction. *Sinkron*, *7*(4), 2407–2414. https://doi.org/10.33395/sinkron.v7i4.11792

Yunitasari, Y., Musdholifah, A., & Sari, A. K. (2019). Sarcasm Detection for Sentiment Analysis in Indonesian Tweets. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, *13*(1), 53. https://doi.org/10.22146/ijccs.41136

Yusupa, A., & Tarigan, V. (n.d.). *JIP (Jurnal Informatika Polinema) Perbandingan Algoritma Maching Learning Dalam Analisis Sentimen Mobil Listrik Di Indonesia Pada Media Sosial Twitter/X*.